AJUSTE DE UM MODELO SARIMA À PRODUÇÃO MENSAL DE AUTOMÓVEIS DE PASSEIO NO BRASIL

Ana Julia Righetto[1]
Luiz Ricardo Nakamura[1]
Manoel Ivanildo Silvestre Bezerra[2]

[1]*Programa de Pós-Graduação em Estatística e Experimentação Agronômica – ESALQ/USP*
[2]*Departamento de Matemática, Estatística e Computação – FCT/UNESP*
{ajrighetto, lrnakamura}@usp.br; manoel@fct.unesp.br

1 INTRODUÇÃO

Em estatística, uma série temporal é todo e qualquer conjunto de observações ordenadas sequencialmente ao longo do tempo, onde a ordem destas observações é fundamental para o estudo. Uma característica de suma importância deste tipo de dados é que as observações vizinhas são dependentes e o interesse é analisar e modelar esta dependência. Ainda, pode-se subdividir a análise de séries temporais em dois subgrupos: 1) a análise é feita no domínio temporal e os modelos são paramétricos; 2) a análise é feita no domínio de frequência e são mais utilizados os modelos não paramétricos.

Dentro do primeiro subgrupo, pode-se citar o modelo SARIMA (BOX, JENKINS e REINSEL, 1994), o qual foi utilizado neste trabalho. Os conceitos da análise de séries temporais foram aplicados em uma série relativa a produção de automóveis (por unidade), no período de janeiro de 1998 à junho de 2008, totalizando 126 períodos; e uma previsão dos dados dos períodos de julho de 2008 a dezembro de 2008 foi realizada. Para uma comparação das previsões, os dados referentes a este período também foram coletados.

Os dados da série foram retirados do IPEAdata (Banco de dados Instituto de Pesquisa Econômica Aplicada) e são referentes à carros de passeio/passageiros e de uso misto; não englobam veículos comerciais leves, como caminhonetes de uso misto, utilitários e caminhonetes de carga, nem veículos comerciais pesados como caminhões e ônibus.

2 MODELO SARIMA

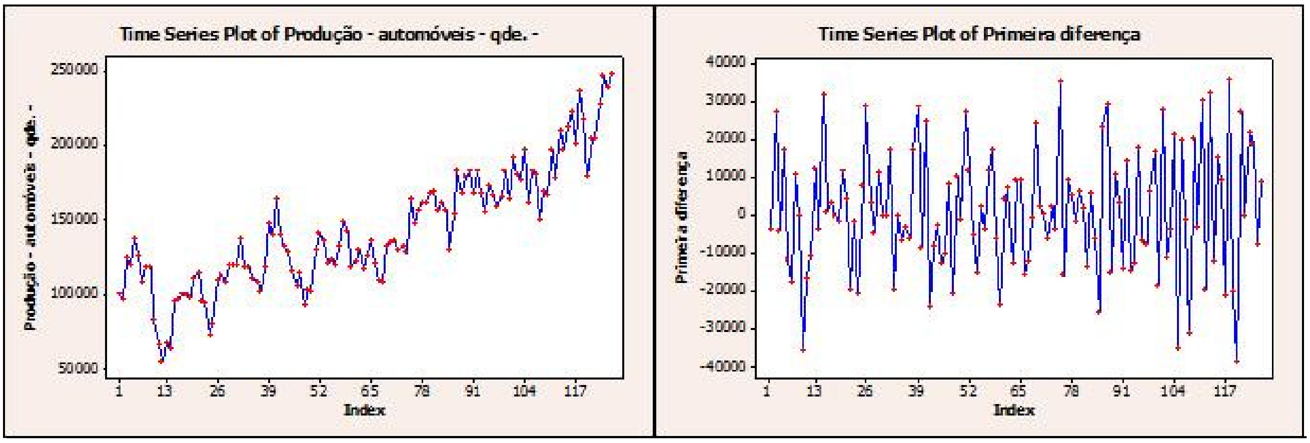
Segundo Morettin e Toloi (2004), mesmo após eliminar a componente sazonal determinística, ainda resta autocorrelação significativa nos *lags* sazonais, ou seja, nos múltiplos de período *s*. Logo, deve-se utilizar um modelo ARIMA sazonal (SARIMA), para considerar a sazonalidade estocástica dos dados. Quando o período s=12, o modelo denominado SARIMA de ordem $(p, d, q) \times (P, D, Q)_{12}$, dado por:

$$\phi(X)\Phi\left(X^{12}\right)\Delta^d \Delta_{12}^D Z_t = \theta(X)\Theta(X)a_t$$

onde a_t é um ruído branco, $\phi(X)$ é o operador autoregressivo (AR) de ordem p, $\theta(X)$ é o operador médias móveis (MA) de ordem q, $\Phi(X)$ é o operador AR-Sazonal de ordem P, $\Theta(X)$ é o operador MA-Sazonal de ordem Q,, Δ^d é o operador diferença e Δ_{12}^D é o operador diferença sazonal.

3 ANÁLISE DOS RESULTADOS

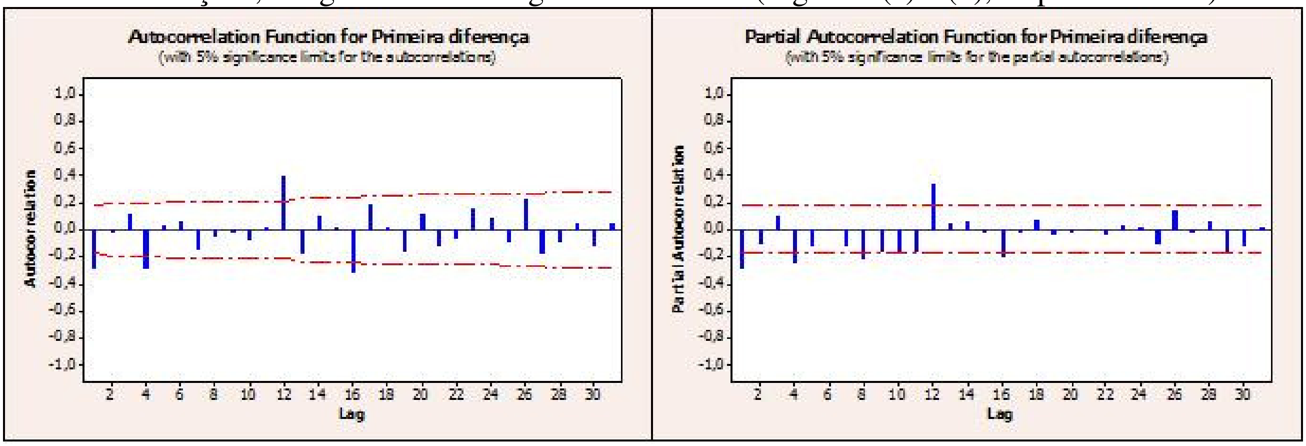
Primeiramente, utilizando o *software* MINITAB, fizemos o gráfico da série temporal mensal da produção de automóveis (por unidade), esboçado na Figura 1(a), e notou-se que a série não era estacionária, sendo assim foi calculada a primeira diferença desta série e um outro gráfico foi esboçado na Figura 1(b), agora referente esta diferença. Todos resultados foram obtidos a partir do MINITAB.



(a) (b)

Figura 1 – (a) Gráfico da série temporal mensal da produção de automóveis (por unidade); (b) Gráfico da primeira diferença da série temporal.

Analisando a Figura 1(b), pode-se notar que a série transformada é estacionária, então se deve agora calcular as autocorrelações amostrais e as autocorrelações parciais amostrais (MORETIN e TOLOI, 2004), da série transformada, para identificação do modelo. Para uma melhor visualização das autocorrelações, a seguir ilustra-se o gráfico de ambas (Figura 2 (a) e (b), respectivamente).



(a) (b)

Figura 2 – (a) Gráfico das autocorrelações amostrais da primeira diferença da série; (b) Gráfico das autorrelações parciais amostrais da primeira diferença da série.

Analisando as Figura 2 (a) e (b), nota-se que existe sazonalidade de período 12. Ainda, é possível observar partes autorregressivas e partes com médias móveis não-sazonais. Portanto, dois modelos foram escolhidos para o ajuste desta série temporal:

1) SARIMA(3,1,2)(0,1,1)$_{12}$ sem o termo constante;
2) SARIMA(3,1,2)(0,2,1)$_{12}$ sem o termo constante.

A seguir faremos uma análise individual de cada um dos modelos identificados.

3.1 Modelo SARIMA(3,1,2)(0,1,1)$_{12}$ sem o termo constante

Com a escolha do modelo, calculou-se a estimação dos parâmetros (Tabela 1) e foi realizado o teste de Box-Pierce (MORETTIN e TOLOI) (Tabela 2) para verificar se o ruído é branco.

Tabela 1 - Estimação dos parâmetros para o modelo SARIMA(3,1,2)(0,1,1)$_{12}$ sem constante

Tipo	Coef	SE Coef	T	P-valor
AR(1)	-1,4726	0,1065	-13,83	0,000
AR(2)	-1,3677	0,1210	-11,31	0,000
AR(3)	-0,3130	0,1041	-3,01	0,003
MA(1)	-1,1409	0,0520	-21,94	0,000

MA(2)	-0,9664	0,0267	-36,21	0,000
SMA	0,8223	0,0801	10,27	0,000

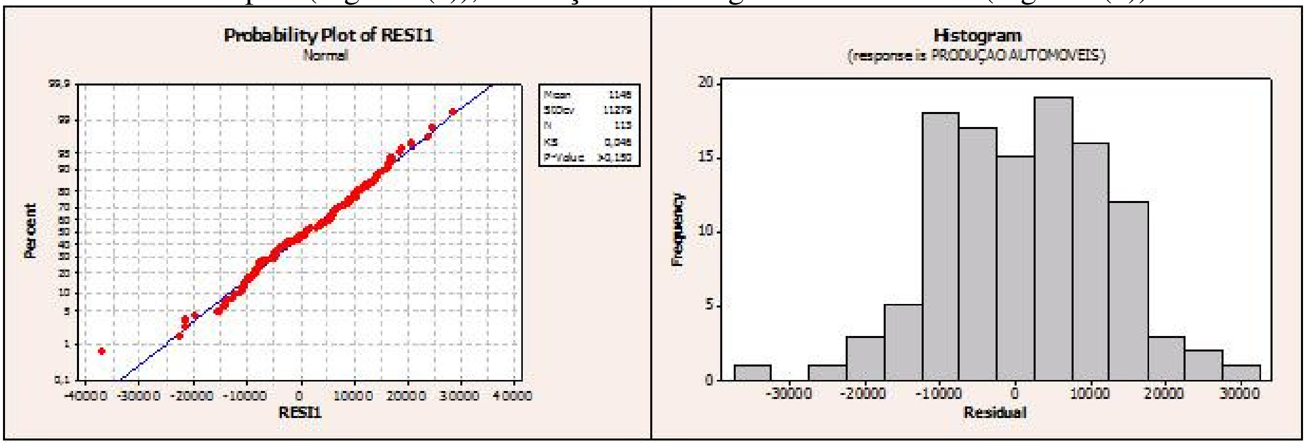
A variância foi obtida no MINITAB utilizando o modulo para análise de Séries Temporais, ou seja, a soma do Quadrado Médio do Erro (MS), foi de MS = 134.555.787 com 107 graus de liberdade, além disso, todos os parâmetros são significativos a um nível de significância de 5%.

Tabela 2 – Teste de Pierce para o modelo SARIMA(3,1,2)(0,1,1)$_{12}$ sem constante

Lag	**12**	**24**	**36**	**48**
χ^2	9,5	21,4	29,6	41,5
g.l	6	18	30	42
P-valor	0,148	0,259	0,488	0,495

Como todos os p-valores da Tabela 2 são maiores que 0,05, pode-se dizer que a um nível de significância de 5% (H_0: O ruído é branco) não rejeita-se a hipótese nula, portanto o ruído é branco (MORETTIN e TOLOI, 2004).

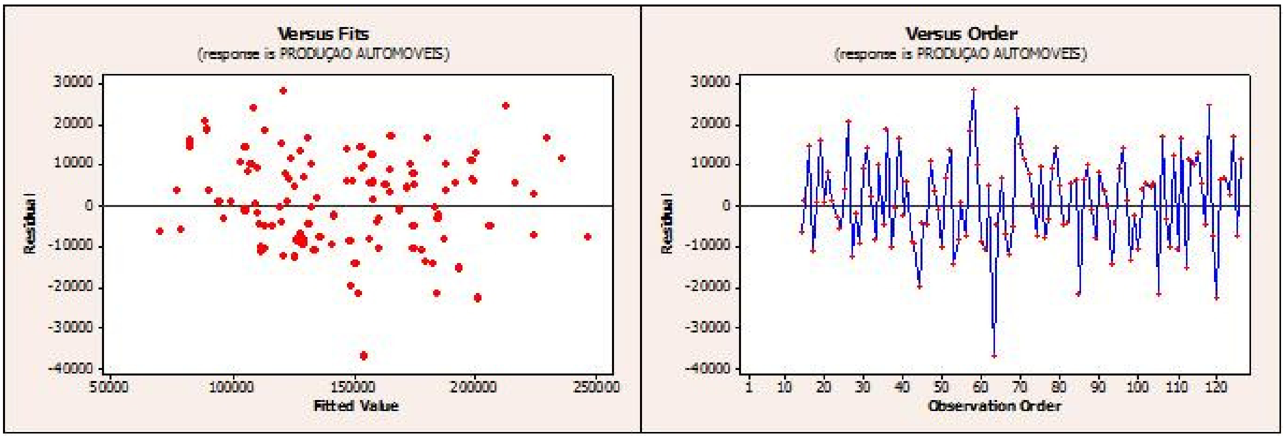
Sendo o ruído branco, o próximo passo agora é fazer a análise dos resíduos. Primeiro testa-se a normalidade e depois (Figura 3(a)), e esboça-se o histograma dos resíduos (Figura 3(b)).



(a) **(b)**

Figura 3 – (a) Teste de normalidade (Kolmogorov-Smirnov, (CONOVER,1999)) dos resíduos; (b) Histograma dos resíduos.

O p-valor retornado no teste de normalidade é >0,15, portanto, a um nível de significância de 5%, não rejeita-se a hipótese nula (H_0: Os resíduos tem distribuição normal), logo os resíduos tem distribuição normal. Com a suposição de normalidade confirmada, estuda-se agora a aleatoriedade dos resíduos:

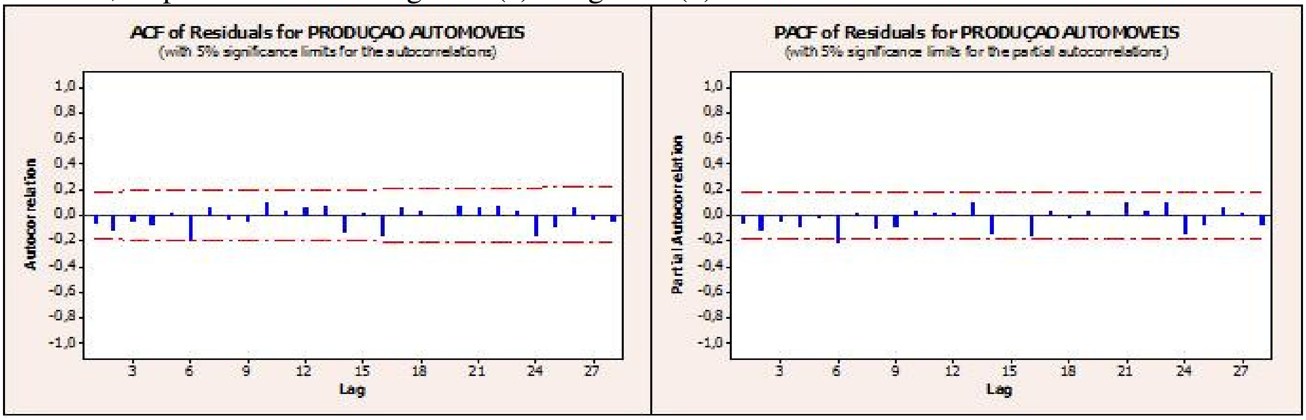


(a) **(b)**

Figura 4 – (a) Gráfico dos resíduos por valores ajustados do modelo; (b) Gráfico de resíduos *versus* ordem.

Nota-se pela Figura 4 (a) e (b) que os resíduos são aleatórios e existe homocedasticidade.

O próximo passo agora é analisar os gráficos de autocorrelação e autocorrelação parcial dos resíduos, respectivamente na Figura 5 (a) e Figura 5 (b).



(a) **(b)**

Figura 5 – (a) Função de autocorrelação dos resíduos; (b) Função de autocorrelação parcial dos resíduos.

Observando a Figura 5 (a) e (b), nota-se que não há somente um valor ligeiramente discrepante, mas que não afetou o teste do Box-Pierce.

Pode-se dizer que depois de todas as análises o ajuste do modelo SARIMA$(3,1,2)(0,1,1)_{12}$ sem o termo constante, é aceitável.

Utilizando novamente o MINITAB, calcularam-se então, as previsões com este modelo e seus respectivos intervalos de 95% de confiança (Tabela 3).

Tabela 3 - Intervalo de 95% de confiança e Previsões do modelo SARIMA$(3,1,2)(0,1,1)_{12}$ sem constante.

Período	Previsão	Limite Inferior	Limite Superior
Julho/2008	251096	228356	273836
Agosto/2008	247871	220520	275222
Setembro/2008	243106	210808	275405
Outubro/2008	255346	218630	292063
Novembro/2008	238351	198017	278685
Dezembro/2008	184592	228292	271991

3.2 Modelo SARIMA$(3,1,2)(0,2,1)_{12}$ sem o termo constante

Para este modelo calculou-se a estimação dos parâmetros (Tabela 4) e foi realizado o teste de Box-Pierce (Tabela 5) para verificar se o ruído é branco.

Tabela 4 – Estimação dos parâmetros SARIMA$(3,1,2)(0,2,1)_{12}$ sem constante.

Tipo	Coef	SE Coef	T	P-valor
AR(1)	-1,9791	0,1363	-14,52	0,000
AR(2)	-1,6574	0,1676	-9,89	0,000
AR(3)	-0,4883	0,1009	-4,84	0,000
MA(1)	-1,6030	0,1036	-15,47	0,000
MA(2)	-0,9152	0,0685	-13,36	0,000
SMA	0,9054	0,0863	10,49	0,000

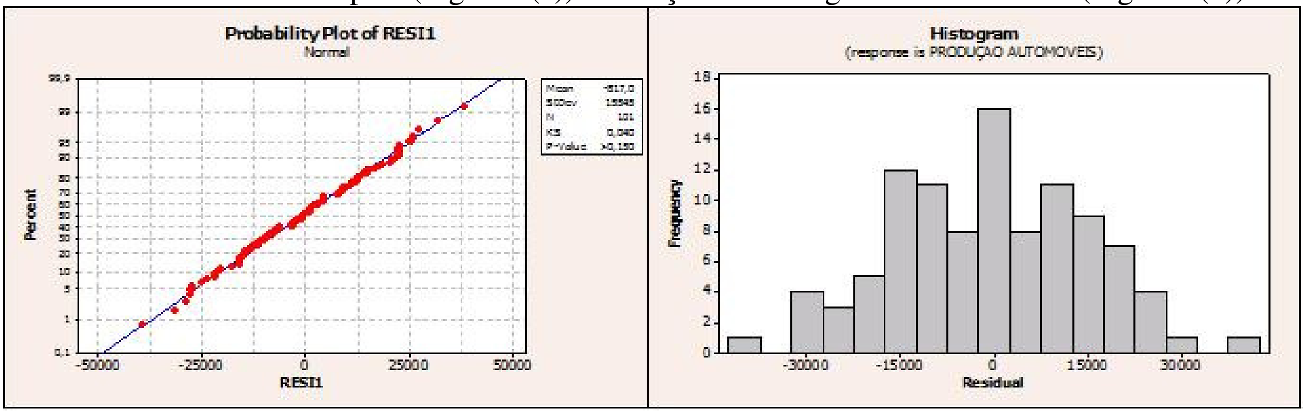
Pode-se notar que todos os parâmetros são significativos a um nível de 5% de significância. Além disso, a variância obtida no MINITAB, ou seja, a soma do Quadrado Médio do Erro (MS), foi igual a 255.080.293 com 95 graus de liberdade.

Tabela 5 – Teste Box-Pierce para o modelo SARIMA(3,1,2)(0,2,1)$_{12}$ sem constante.

Lag	**12**	**24**	**36**	**48**
χ^2	7,8	27,6	36,5	50,3
g.l	6	18	30	42
P-valor	0,256	0,069	0,193	0,18

Observa-se pela Tabela 5, que todos os p-valores são maiores que 5%, portanto, a um nível de significância de 5% não rejeita-se a hipótese de nulidade (H_0: O ruído é branco), logo o ruído é branco.

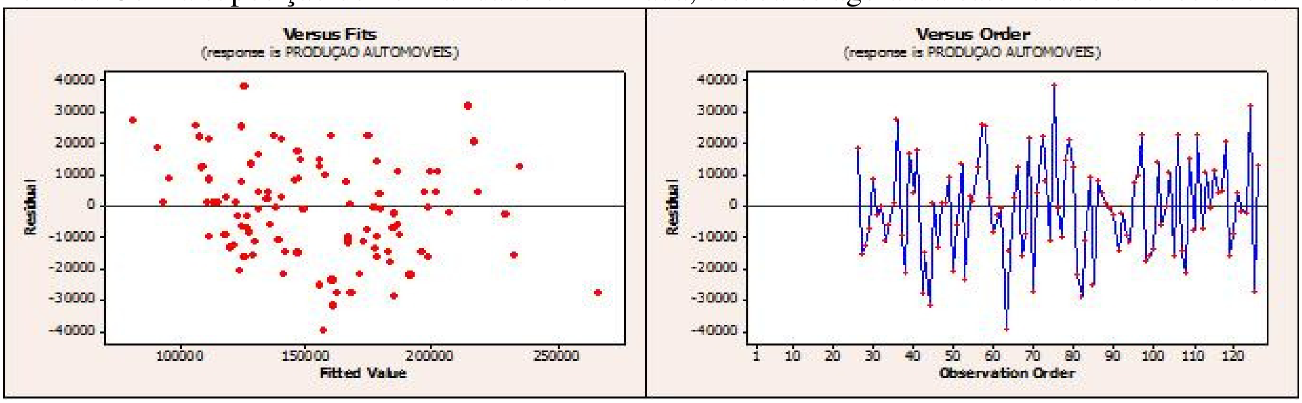
Como se tem o ruído branco, o próximo passo agora é fazer a análise dos resíduos. Portanto, testa-se a normalidade e depois (Figura 6(a)) e esboça-se o histograma dos resíduos (Figura 6(b)).



(a) **(b)**

Figura 6 – (a) Teste de normalidade (Kolmogorov-Smirnov, (CONOVER, 1999)) dos resíduos; (b) Histograma dos resíduos

O p-valor retornado no teste de normalidade é >0,15, portanto, a um nível de significância de 5%, não rejeita-se a hipótese nula (H_0: A distribuição é normal), logo os resíduos tem distribuição normal. Com a suposição de normalidade confirmada, estuda-se agora a aleatoriedade dos resíduos:

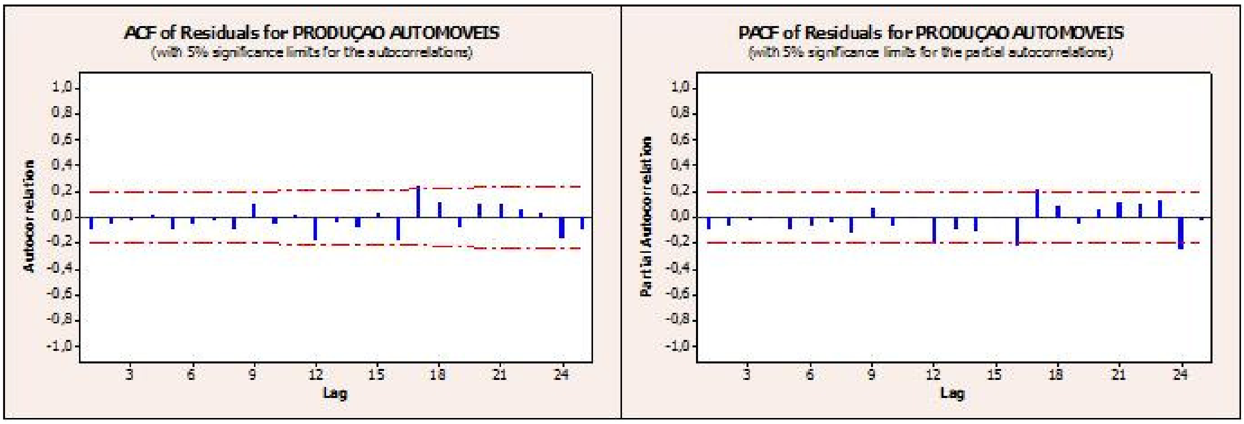


(a) **(b)**

Figura 7 – (a) Gráfico dos resíduos por valores ajustados do modelo; (b) Gráfico de resíduos *versus* ordem.

Nota-se pela Figura 7 (a) e (b) que os resíduos são aleatórios e existe homocedasticidade.

O próximo passo agora é analisar os gráficos de autocorrelação e autocorrelação parcial dos resíduos, respectivamente na Figura 8 (a) e Figura 8 (b).



(a) **(b)**

Figura 8 – (a) Função de autocorrelação dos resíduos; (b) Função de autocorrelação parcial dos resíduos.

Observando a Figura 8 (a) e (b) nota-se que não há alguns valores discrepantes, mas que não afetaram o teste do Box-Pierce.

Pode-se dizer que depois de todas as análises o ajuste do modelo SARIMA(3,1,2)(0,2,1)$_{12}$ sem o termo constante, é também aceitável. Utilizando novamente o MINITAB, foram obtidas, por meio deste modelo, as previsões e seus respectivos intervalos de 95% de confiança (Tabela 6).

Tabela 6 - Intervalo de 95% de confiança e Previsões do modelo SARIMA(3,1,2)(0,1,1)$_{12}$ sem constante.

Período	Previsão	Limite Inferior	Limite Superior
Julho/2008	270593	239284	301903
Agosto/2008	269078	232174	305982
Setembro/2008	249335	207548	291122
Outubro/2008	301426	253389	349464
Novembro/2008	262973	210452	315494
Dezembro/2008	235839	179888	291790

4 CONCLUSÃO

Os dois modelos mostram-se adequados, uma vez que quase todas as suposições necessárias para o ajuste dos dados foram aceitas. Como critério de decisão para escolha do melhor modelo ajustado, utilizou-se o critério do Erro Quadrático Médio Mínimo de Precisão (EQMMP) (MORETTIN e TOLOI, 2004). No caso, quanto menor o valor do EQMMP, mais adequado está o modelo. Na tabela a seguir (Tabela 7) têm-se os EQMMP dos dois modelos:

Tabela 7 – Valores dos EQMMP dos modelos ajustados.

MODELO	EQMMP
SARIMA(3,1,2)(0,1,1)$_{12}$	5200819596,5
SARIMA(3,1,2)(0,2,1)$_{12}$	7288248255,0

Portanto, pelo critério do EQMMP, pode-se escolher o modelo SARIMA(3,1,2)(0,1,1)$_{12}$, para fazer previsões da série referente a produção de automóveis no Brasil (por unidade), pois o este modelo forneceu o menor EQMMP.

REFERÊNCIAS

BOX, G.E.P.; JENKINS,G.W. and REINSEL, G.C. *Time Series Analysis, Forecasting and Control.* Third Edition. Prentice-Hall, Englewood Cliffs, N.J., 1994.

BRASIL. Instituto de Pesquisa Econômica Aplicada (IPEA). Disponível em: http://www.ipeadata.gov.br. Acesso em: 13 de novembro de 2009.

CONOVER, W.J. *Pratical Nonparametric Statistics.* Third Edition. John Wiley & Sons, N.Y., 1999.

MORETTIN, P.A.; TOLOI, C.M.C. Análise de séries temporais. 2ª ed. São Paulo: Editora Edgarg Blucher, 2004.